

## SEQUENCE LISTING

# 6

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: BASF Aktiengesellschaft
- (B) STREET: Karl Bosch Strasse
- (C) CITY: Ludwigshafen
- (D) FEDERAL STATE: Rheinland-Pfalz
- (E) COUNTRY: Germany
- (F) POSTAL CODE: 67056

(ii) TITLE OF APPLICATION: Process for preparing biotin

(iii) NUMBER OF SEQUENCES: 15

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

## (2) INFORMATION FOR SEQ ID No: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1155 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

## (vi) ORIGINAL SOURCE:

- (B) STRAIN: Escherichia coli

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: metK

## (ix) FEATURES:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1155

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 1:

ATG	GCA	AAA	CAC	CTT	TTT	ACG	TCC	GAG	TCC	GTC	TCT	GAA	GGG	CAT	CCT	48
Met	Ala	Lys	His	Leu	Phe	Thr	Ser	Glu	Ser	Val	Ser	Glu	Gly	His	Pro	
1				5				10						15		
GAC	AAA	ATT	GCT	GAC	CAA	ATT	TCT	GAT	GCC	GTT	TTA	GAC	GCG	ATC	CTC	96
Asp	Lys	Ile	Ala	Asp	Gln	Ile	Ser	Asp	Ala	Val	Leu	Asp	Ala	Ile	Leu	
			20					25					30			
GAA	CAG	GAT	CCG	AAA	GCA	CGC	GTT	GCT	TGC	GAA	ACC	TAC	GTA	AAA	ACC	144
Glu	Gln	Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr	
		35					40					45				
GGC	ATG	GTT	TTA	GTT	GGC	GGC	GAA	ATC	ACC	ACC	AGC	GCC	TGG	GTA	GAC	192
Gly	Met	Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp	
	50					55					60					
ATC	GAA	GAG	ATC	ACC	CGT	AAC	ACC	GTT	CGC	GAA	ATT	GGC	TAT	GTG	CAT	240
Ile	Glu	Glu	Ile	Thr	Arg	Asn	Thr	Val	Arg	Glu	Ile	Gly	Tyr	Val	His	
65					70					75					80	
TCC	GAC	ATG	GGC	TTT	GAC	GCT	AAC	TCC	TGT	GCG	GTT	CTG	AGC	GCT	ATC	288
Ser	Asp	Met	Gly	Phe	Asp	Ala	Asn	Ser	Cys	Ala	Val	Leu	Ser	Ala	Ile	
				85					90					95		
GGC	AAA	CAG	TCT	CCT	GAC	ATC	AAC	CAG	GGC	GTT	GAC	CGT	GCC	GAT	CCG	336
Gly	Lys	Gln	Ser	Pro	Asp	Ile	Asn	Gln	Gly	Val	Asp	Arg	Ala	Asp	Pro	
			100					105					110			
CTG	GAA	CAG	GGC	GCG	GGT	GAC	CAG	GGT	CTG	ATG	TTT	GGC	TAC	GCA	ACT	384
Leu	Glu	Gln	Gly	Ala	Gly	Asp	Gln	Gly	Leu	Met	Phe	Gly	Tyr	Ala	Thr	
		115					120					125				
AAT	GAA	ACC	GAC	GTG	CTG	ATG	CCA	GCA	CCT	ATC	ACC	TAT	GCA	CAC	CGT	432
Asn	Glu	Thr	Asp	Val	Leu	Met	Pro	Ala	Pro	Ile	Thr	Tyr	Ala	His	Arg	
	130						135					140				
CTG	GTA	CAG	CGT	CAG	GCT	GAA	GTG	CGT	AAA	AAC	GGC	ACT	CTG	CCG	TGG	480
Leu	Val	Gln	Arg	Gln	Ala	Glu	Val	Arg	Lys	Asn	Gly	Thr	Leu	Pro	Trp	
145					150					155					160	
CTG	CGC	CCG	GAC	GCG	AAA	AGC	CAG	GTG	ACT	TTT	CAG	TAT	GAC	GAC	GGC	528
Leu	Arg	Pro	Asp	Ala	Lys	Ser	Gln	Val	Thr	Phe	Gln	Tyr	Asp	Asp	Gly	
				165					170					175		
AAA	ATC	GTT	GGT	ATC	GAT	GCT	GTC	GTG	CTT	TCC	ACT	CAG	CAC	TCT	GAA	536
Lys	Ile	Val	Gly	Ile	Asp	Ala	Val	Val	Leu	Ser	Thr	Gln	His	Ser	Glu	
			180					185						190		

GAG ATC GAC CAG AAA TCG CTG CAA GAA GCG GTA ATG GAA GAG ATC ATC	624
Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile	
195 200 205	
AAG CCA ATT CTG CCC GCT GAA TGG CTG ACT TCT GCC ACC AAA TTC TTC	672
Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe	
210 215 220	
ATC AAC CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC	720
Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys	
225 230 235 240	
GGT CTG ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG	768
Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala	
245 250 255	
CGT CAC GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC	816
Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp	
260 265 270	
CGT TCC GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT	864
Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala	
275 280 285	
GCT GGC CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC	912
Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly	
290 295 300	
GTG GCT GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA	960
Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys	
305 310 315 320	
GTG CCT TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG	1008
Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu	
325 330 335	
CGC CCA TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC	1056
Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr	
340 345 350	
AAA GAA ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG	1104
Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp	
355 360 365	
GAA AAA ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG	1152
Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys	
370 375 380	

TAA

1155

385

## (2) INFORMATION FOR SEQ ID No: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 384 Amino acids

(B) TYPE: Amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 2:

Met	Ala	Lys	His	Leu	Phe	Thr	Ser	Glu	Ser	Val	Ser	Glu	Gly	His	Pro	1	5	10	15
Asp	Lys	Ile	Ala	Asp	Gln	Ile	Ser	Asp	Ala	Val	Leu	Asp	Ala	Ile	Leu	20	25	30	
Glu	Gln	Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr	35	40	45	
Gly	Met	Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp	50	55	60	
Ile	Glu	Glu	Ile	Thr	Arg	Asn	Thr	Val	Arg	Glu	Ile	Gly	Tyr	Val	His	65	70	75	80
Ser	Asp	Met	Gly	Phe	Asp	Ala	Asn	Ser	Cys	Ala	Val	Leu	Ser	Ala	Ile	85	90	95	
Gly	Lys	Gln	Ser	Pro	Asp	Ile	Asn	Gln	Gly	Val	Asp	Arg	Ala	Asp	Pro	100	105	110	
Leu	Glu	Gln	Gly	Ala	Gly	Asp	Gln	Gly	Leu	Met	Phe	Gly	Tyr	Ala	Thr	115	120	125	
Asn	Glu	Thr	Asp	Val	Leu	Met	Pro	Ala	Pro	Ile	Thr	Tyr	Ala	His	Arg	130	135	140	
Leu	Val	Gln	Arg	Gln	Ala	Glu	Val	Arg	Lys	Asn	Gly	Thr	Leu	Pro	Trp	145	150	155	160
Leu	Arg	Pro	Asp	Ala	Lys	Ser	Gln	Val	Thr	Phe	Gln	Tyr	Asp	Asp	Gly	165	170	175	
Lys	Ile	Val	Gly	Ile	Asp	Ala	Val	Val	Leu	Ser	Thr	Gln	His	Ser	Glu	180	185	190	

Glu	Ile	Asp	Gln	Lys	Ser	Leu	Gln	Glu	Ala	Val	Met	Glu	Glu	Ile	Ile	195	200	205	
Lys	Pro	Ile	Leu	Pro	Ala	Glu	Trp	Leu	Thr	Ser	Ala	Thr	Lys	Phe	Phe	210	215	220	
Ile	Asn	Pro	Thr	Gly	Arg	Phe	Val	Ile	Gly	Gly	Pro	Met	Gly	Asp	Cys	225	230	235	240
Gly	Leu	Thr	Gly	Arg	Lys	Ile	Ile	Val	Asp	Thr	Tyr	Gly	Gly	Met	Ala	245	250	255	
Arg	His	Gly	Gly	Gly	Ala	Phe	Ser	Gly	Lys	Asp	Pro	Ser	Lys	Val	Asp	260	265	270	
Arg	Ser	Ala	Ala	Tyr	Ala	Ala	Arg	Tyr	Val	Ala	Lys	Asn	Ile	Val	Ala	275	280	285	
Ala	Gly	Leu	Ala	Asp	Arg	Cys	Glu	Ile	Gln	Val	Ser	Tyr	Ala	Ile	Gly	290	295	300	
Val	Ala	Glu	Pro	Thr	Ser	Ile	Met	Val	Glu	Thr	Phe	Gly	Thr	Glu	Lys	305	310	315	320
Val	Pro	Ser	Glu	Gln	Leu	Thr	Leu	Leu	Val	Arg	Glu	Phe	Phe	Asp	Leu	325	330	335	
Arg	Pro	Tyr	Gly	Leu	Ile	Gln	Met	Leu	Asp	Leu	Leu	His	Pro	Ile	Tyr	340	345	350	
Lys	Glu	Thr	Ala	Ala	Tyr	Gly	His	Phe	Gly	Arg	Glu	His	Phe	Pro	Trp	355	360	365	
Glu	Lys	Thr	Asp	Lys	Ala	Gln	Leu	Leu	Arg	Asp	Ala	Ala	Gly	Leu	Lys	370	375	380	

## (2) INFORMATION FOR SEQ ID No: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1206 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

## (vi) ORIGINAL SOURCE:

(B) STRAIN: Escherichia coli

## (vii) IMMEDIATE SOURCE:

(B) CLONE: bioS1

## (ix) FEATURES:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1206

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 3:

ATG AAC GTT TTT AAT CCC GCG CAG TTT CGC GCC CAG TTT CCC GCA CTA	48
Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu	
1 5 10 15	
CAG GAT GCG GGC GTC TAT CTC GAC AGC GCC GCG ACC GCG CTT AAA CCT	96
Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro	
20 25 30	
GAA GCC GTG GTT GAA GCC ACC CAA CAG TTT TAC AGT CTG AGC GCC GGA	144
Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly	
35 40 45	
AAC GTC CAT CGC AGC CAG TTT GCC GAA GCC CAA CGC CTG ACC GCG CGT	192
Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg	
50 55 60	
TAT GAA GCT GCA CGA GAG AAA GTG GCG CAA TTA CTG AAT GCA CCG GAT	240
Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp	
65 70 75 80	
GAT AAA ACT ATC GTC TGG ACG CGC GGC ACC ACT GAA TCC ATC AAC ATG	288
Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met	
85 90 95	
GTG GCA CAA TGC TAT GCG CGT CCG CGT CTG CAA CCG GGC GAT GAG ATT	336
Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile	
100 105 110	
ATT GTC AGC GTG GCA GAA CAC CAC GCC AAC CTC GTC CCC TGG CTG ATG	384
Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met	
115 120 125	
GTC GCC CAA CAA ACT GGA GCC AAA GTG GTG AAA TTG CCG CTT AAT GCG	432
Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala	
130 135 140	

CAG	CGA	CTG	CCG	GAT	GTC	GAT	TTG	TTG	CCA	GAA	CTG	ATT	ACT	CCC	CGT	480
Gln	Arg	Leu	Pro	Asp	Val	Asp	Leu	Leu	Pro	Glu	Leu	Ile	Thr	Pro	Arg	
145					150					155					160	
AGT	CGG	ATT	CTG	GCG	TTG	GGT	CAG	ATG	TCG	AAC	GTT	ACT	GGC	GGT	TGC	528
Ser	Arg	Ile	Leu	Ala	Leu	Gly	Gln	Met	Ser	Asn	Val	Thr	Gly	Gly	Cys	
				165					170					175		
CCG	GAT	CTG	GCG	CGA	GCG	ATT	ACC	TTT	GCT	CAT	TCA	GCC	GGG	ATG	GTG	576
Pro	Asp	Leu	Ala	Arg	Ala	Ile	Thr	Phe	Ala	His	Ser	Ala	Gly	Met	Val	
			180					185					190			
GTG	ATG	GTT	GAT	GGT	GCT	CAG	GGG	GCA	GTG	CAT	TTC	CCC	GCG	GAT	GTT	624
Val	Met	Val	Asp	Gly	Ala	Gln	Gly	Ala	Val	His	Phe	Pro	Ala	Asp	Val	
	195					200						205				
CAG	CAA	CTG	GAT	ATT	GAT	TTC	TAT	GCT	TTT	TCA	GGT	CAC	AAA	CTG	TAT	672
Gln	Gln	Leu	Asp	Ile	Asp	Phe	Tyr	Ala	Phe	Ser	Gly	His	Lys	Leu	Tyr	
	210					215					220					
GGC	CCG	ACA	GGT	ATC	GGC	GTG	CTG	TAT	GGT	AAA	TCA	GAA	CTG	CTG	GAG	720
Gly	Pro	Thr	Gly	Ile	Gly	Val	Leu	Tyr	Gly	Lys	Ser	Glu	Leu	Leu	Glu	
225					230					235					240	
GCG	ATG	TCG	CCC	TGG	CTG	GGC	GGC	GGC	AAA	ATG	GTT	CAC	GAA	GTG	AGT	768
Ala	Met	Ser	Pro	Trp	Leu	Gly	Gly	Gly	Lys	Met	Val	His	Glu	Val	Ser	
				245					250					255		
TTT	GAC	GGC	TTC	ACG	ACT	CAA	TCT	GCG	CCG	TGG	AAA	CTG	GAA	GCT	GGA	816
Phe	Asp	Gly	Phe	Thr	Thr	Gln	Ser	Ala	Pro	Trp	Lys	Leu	Glu	Ala	Gly	
			260					265					270			
ACG	CCA	AAT	GTC	GCT	GGT	GTC	ATA	GGA	TTA	AGC	GCG	GCG	CTG	GAA	TGG	864
Thr	Pro	Asn	Val	Ala	Gly	Val	Ile	Gly	Leu	Ser	Ala	Ala	Leu	Glu	Trp	
		275					280					285				
CTG	GCA	GAT	TAC	GAT	ATC	AAC	CAG	GCC	GAA	AGC	TGG	AGC	CGT	AGC	TTA	912
Leu	Ala	Asp	Tyr	Asp	Ile	Asn	Gln	Ala	Glu	Ser	Trp	Ser	Arg	Ser	Leu	
	290					295					300					
GCA	ACG	CTG	GCG	GAA	GAT	GCG	CTG	GCG	AAA	CGT	CCC	GGC	TTT	CGT	TCA	960
Ala	Thr	Leu	Ala	Glu	Asp	Ala	Leu	Ala	Lys	Arg	Pro	Gly	Phe	Arg	Ser	
305					310					315					320	
TTT	CGC	TGC	CAG	GAT	TCC	AGC	CTG	CTG	GCC	TTT	GAT	TTT	GCT	GGC	GTT	1008
Phe	Arg	Cys	Gln	Asp	Ser	Ser	Leu	Leu	Ala	Phe	Asp	Phe	Ala	Gly	Val	
			325						330					335		

CAT CAT AGC GAT ATG GTG ACG CTG CTG GCG GAG TAC GGT ATT GCC CTG	1056
His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu	
340 345 350	
CGG GCC GGG CAG CAT TGC GCT CAG CCG CTA CTG GCA GAA TTA GGC GTA	1104
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val	
355 360 365	
ACC GGC ACA CTG CGC GCC TCT TTT GCG CCA TAT AAT ACA AAG AGT GAT	1152
Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp	
370 375 380	
GTG GAT GCG CTG GTG AAT GCC GTT GAC CGC GCG CTG GAA TTA TTG GTG	1200
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val	
385 390 395 400	
GAT TAA	1206
Asp	

## (2) INFORMATION FOR SEQ ID No: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 401 Amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 4:

Met	Asn	Val	Phe	Asn	Pro	Ala	Gln	Phe	Arg	Ala	Gln	Phe	Pro	Ala	Leu
1				5				10					15		
Gln	Asp	Ala	Gly	Val	Tyr	Leu	Asp	Ser	Ala	Ala	Thr	Ala	Leu	Lys	Pro
		20						25					30		
Glu	Ala	Val	Val	Glu	Ala	Thr	Gln	Gln	Phe	Tyr	Ser	Leu	Ser	Ala	Gly
		35					40					45			
Asn	Val	His	Arg	Ser	Gln	Phe	Ala	Glu	Ala	Gln	Arg	Leu	Thr	Ala	Arg
	50					55					60				
<del>Tyr</del>	<del>Glu</del>	<del>Ala</del>	<del>Ala</del>	<del>Arg</del>	<del>Glu</del>	<del>Lys</del>	<del>Val</del>	<del>Ala</del>	<del>Gln</del>	<del>Leu</del>	<del>Leu</del>	<del>Asn</del>	<del>Ala</del>	<del>Pro</del>	<del>Asp</del>
65						70				75				80	
Asp	Lys	Thr	Ile	Val	Trp	Thr	Arg	Gly	Thr	Thr	Glu	Ser	Ile	Asn	Met
				85					90					95	



Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile  
100 105 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met  
115 120 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala  
130 135 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg  
145 150 155 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys  
165 170 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val  
180 185 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val  
195 200 205

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr  
210 215 220

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu  
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser  
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly  
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp  
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu  
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser  
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val  
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu  
340 345 350

Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val  
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp  
 370 375 380

Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val  
 385 390 395 400

Asp

(2) INFORMATION FOR SEQ ID No: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1215 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vi) ORIGINAL SOURCE:

- (B) STRAIN: Escherichia coli

(vii) IMMEDIATE SOURCE:

- (B) CLONE: bioS2

(ix) FEATURES:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1215

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 5:

ATG AAA TTA CCG ATT TAT CTC GAC TAC TCC GCA ACC ACG CCG GTG GAC	48
Met Lys Leu Pro Ile Tyr Leu Asp Tyr Ser Ala Thr Thr Pro Val Asp	
1 5 10 15	
CCG CGT GTT GCC GAG AAA ATG ATG CAG TTT ATG ACG ATG GAC GGA ACC	96
Pro Arg Val Ala Glu Lys Met Met Gln Phe Met Thr Met Asp Gly Thr	
20 25 30	
<del>TTT GGT AAC CCG GCC TCC CGT TCT CAC CGT TTC GGC TGG CAG GCT GAA</del>	<del>144</del>
<del>Phe Gly Asn Pro Ala Ser Arg Ser His Arg Phe Gly Trp Gln Ala Glu</del>	
<del>35 40 45</del>	
GAA GCG GTA GAT ATC GCC CGT AAT CAG ATT GCC GAT CTG GTC GGC GCT	192
Glu Ala Val Asp Ile Ala Arg Asn Gln Ile Ala Asp Leu Val Gly Ala	
50 55 60	

GAT	CCG	CGT	GAA	ATC	GTC	TTT	ACC	TCT	GGT	GCA	ACC	GAA	TCT	GAC	AAC	240
Asp	Pro	Arg	Glu	Ile	Val	Phe	Thr	Ser	Gly	Ala	Thr	Glu	Ser	Asp	Asn	
65					70					75					80	
CTG	GCG	ATC	AAA	GGT	GCA	GCC	AAC	TTT	TAT	CAG	AAA	AAA	GGC	AAG	CAC	288
Leu	Ala	Ile	Lys	Gly	Ala	Ala	Asn	Phe	Tyr	Gln	Lys	Lys	Gly	Lys	His	
				85					90					95		
ATC	ATC	ACC	AGC	AAA	ACC	GAA	CAC	AAA	GCG	GTA	CTG	GAT	ACC	TGC	CGT	336
Ile	Ile	Thr	Ser	Lys	Thr	Glu	His	Lys	Ala	Val	Leu	Asp	Thr	Cys	Arg	
			100					105					110			
CAG	CTG	GAG	CGC	GAA	GGT	TTT	GAA	GTC	ACC	TAC	CTG	GCA	CCG	CAG	CGT	384
Gln	Leu	Glu	Arg	Glu	Gly	Phe	Glu	Val	Thr	Tyr	Leu	Ala	Pro	Gln	Arg	
		115					120					125				
AAC	GGC	ATT	ATC	GAC	CTG	AAA	GAA	CTT	GAA	GCA	GCG	ATG	CGT	GAC	GAC	432
Asn	Gly	Ile	Ile	Asp	Leu	Lys	Glu	Leu	Glu	Ala	Ala	Met	Arg	Asp	Asp	
	130					135					140					
ACC	ATC	CTC	GTG	TCC	ATC	ATG	CAC	GTA	AAT	AAC	GAA	ATC	GGC	GTG	GTG	480
Thr	Ile	Leu	Val	Ser	Ile	Met	His	Val	Asn	Asn	Glu	Ile	Gly	Val	Val	
145					150					155					160	
CAG	GAT	ATC	GCG	GCT	ATC	GGC	GAA	ATG	TGC	CGT	GCT	CGT	GGC	ATT	ATC	528
Gln	Asp	Ile	Ala	Ala	Ile	Gly	Glu	Met	Cys	Arg	Ala	Arg	Gly	Ile	Ile	
				165					170					175		
TAT	CAC	GTT	GAT	GCA	ACC	CAG	AGC	GTG	GGT	AAA	CTG	CCT	ATC	GAC	CTG	576
Tyr	His	Val	Asp	Ala	Thr	Gln	Ser	Val	Gly	Lys	Leu	Pro	Ile	Asp	Leu	
			180					185					190			
AGC	CAG	TTG	AAA	GTT	GAC	CTG	ATG	TCT	TTC	TCC	GGT	CAC	AAA	ATC	TAT	624
Ser	Gln	Leu	Lys	Val	Asp	Leu	Met	Ser	Phe	Ser	Gly	His	Lys	Ile	Tyr	
		195					200					205				
GGC	CCG	AAA	GGT	ATC	GGT	GCG	CTG	TAT	GTA	CGT	CGT	AAA	CCG	CGC	GTA	672
Gly	Pro	Lys	Gly	Ile	Gly	Ala	Leu	Tyr	Val	Arg	Arg	Lys	Pro	Arg	Val	
	210					215					220					
CGC	ATC	GAA	GCG	CAA	ATG	CAC	GGC	GGC	GGT	CAC	GAG	CGC	GGT	ATG	CGT	720
Arg	Ile	Glu	Ala	Gln	Met	His	Gly	Gly	Gly	His	Glu	Arg	Gly	Met	Arg	
225					230					235					240	
TCC	GGC	ACT	CTG	CCT	GTT	CAC	CAG	ATC	GTC	GGA	ATG	GGC	GAG	GCC	TAT	768
Ser	Gly	Thr	Leu	Pro	Val	His	Gln	Ile	Val	Gly	Met	Gly	Glu	Ala	Tyr	
				245					250					255		

CGC ATC GCA AAA GAA GAG ATG GCG ACC GAG ATG GAA CGT CTG CGC GGC	816
Arg Ile Ala Lys Glu Glu Met Ala Thr Glu Met Glu Arg Leu Arg Gly	
260 265 270	
CTG CGT AAC CGT CTG TGG AAC GGC ATC AAA GAT ATC GAA GAA GTT TAC	864
Leu Arg Asn Arg Leu Trp Asn Gly Ile Lys Asp Ile Glu Glu Val Tyr	
275 280 285	
CTG AAC GGT GAC CTG GAA CAC GGT GCG CCG AAC ATT CTC AAC GTC AGC	912
Leu Asn Gly Asp Leu Glu His Gly Ala Pro Asn Ile Leu Asn Val Ser	
290 295 300	
TTC AAC TAC GTT GAA GGT GAG TCG CTG ATT ATG GCG CTG AAA GAC CTC	960
Phe Asn Tyr Val Glu Gly Glu Ser Leu Ile Met Ala Leu Lys Asp Leu	
305 310 315 320	
GCA GTT TCT TCA GGT TCC GCC TGT ACG TCA GCA AGC CTC GAA CCG TCC	1008
Ala Val Ser Ser Gly Ser Ala Cys Thr Ser Ala Ser Leu Glu Pro Ser	
325 330 335	
TAC GTG CTG CGC GCG CTG GGG CTG AAC GAC GAG CTG GCA CAT AGC TCT	1056
Tyr Val Leu Arg Ala Leu Gly Leu Asn Asp Glu Leu Ala His Ser Ser	
340 345 350	
ATC CGT TTC TCT TTA GGT CGT TTT ACT ACT GAA GAA GAG ATC GAC TAC	1104
Ile Arg Phe Ser Leu Gly Arg Phe Thr Thr Glu Glu Glu Ile Asp Tyr	
355 360 365	
ACC ATC GAG TTA GTT CGT AAA TCC ATC GGT CGT CTG CGT GAC CTT TCT	1152
Thr Ile Glu Leu Val Arg Lys Ser Ile Gly Arg Leu Arg Asp Leu Ser	
370 375 380	
CCG CTG TGG GAA ATG TAC AAG CAG GGC GTG GAT CTG AAC AGC ATC GAA	1200
Pro Leu Trp Glu Met Tyr Lys Gln Gly Val Asp Leu Asn Ser Ile Glu	
385 390 395 400	
TGG GCT CAT CAT TAA	1215
Trp Ala His His	
405	

## (2) INFORMATION FOR SEQ ID No: 6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 Amino acids  
 (B) TYPE: Amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 6:

Met	Lys	Leu	Pro	Ile	Tyr	Leu	Asp	Tyr	Ser	Ala	Thr	Thr	Pro	Val	Asp	1	5	10	15
Pro	Arg	Val	Ala	Glu	Lys	Met	Met	Gln	Phe	Met	Thr	Met	Asp	Gly	Thr	20	25	30	
Phe	Gly	Asn	Pro	Ala	Ser	Arg	Ser	His	Arg	Phe	Gly	Trp	Gln	Ala	Glu	35	40	45	
Glu	Ala	Val	Asp	Ile	Ala	Arg	Asn	Gln	Ile	Ala	Asp	Leu	Val	Gly	Ala	50	55	60	
Asp	Pro	Arg	Glu	Ile	Val	Phe	Thr	Ser	Gly	Ala	Thr	Glu	Ser	Asp	Asn	65	70	75	80
Leu	Ala	Ile	Lys	Gly	Ala	Ala	Asn	Phe	Tyr	Gln	Lys	Lys	Gly	Lys	His	85	90	95	
Ile	Ile	Thr	Ser	Lys	Thr	Glu	His	Lys	Ala	Val	Leu	Asp	Thr	Cys	Arg	100	105	110	
Gln	Leu	Glu	Arg	Glu	Gly	Phe	Glu	Val	Thr	Tyr	Leu	Ala	Pro	Gln	Arg	115	120	125	
Asn	Gly	Ile	Ile	Asp	Leu	Lys	Glu	Leu	Glu	Ala	Ala	Met	Arg	Asp	Asp	130	135	140	
Thr	Ile	Leu	Val	Ser	Ile	Met	His	Val	Asn	Asn	Glu	Ile	Gly	Val	Val	145	150	155	160
Gln	Asp	Ile	Ala	Ala	Ile	Gly	Glu	Met	Cys	Arg	Ala	Arg	Gly	Ile	Ile	165	170	175	
Tyr	His	Val	Asp	Ala	Thr	Gln	Ser	Val	Gly	Lys	Leu	Pro	Ile	Asp	Leu	180	185	190	
Ser	Gln	Leu	Lys	Val	Asp	Leu	Met	Ser	Phe	Ser	Gly	His	Lys	Ile	Tyr	195	200	205	
Gly	Pro	Lys	Gly	Ile	Gly	Ala	Leu	Tyr	Val	Arg	Arg	Lys	Pro	Arg	Val	210	215	220	
Arg	Ile	Glu	Ala	Gln	Met	His	Gly	Gly	Gly	His	Glu	Arg	Gly	Met	Arg	225	230	235	240
Ser	Gly	Thr	Leu	Pro	Val	His	Gln	Ile	Val	Gly	Met	Gly	Glu	Ala	Tyr	245	250	255	

Arg Ile Ala Lys Glu Glu Met Ala Thr Glu Met Glu Arg Leu Arg Gly  
 260 265 270  
 Leu Arg Asn Arg Leu Trp Asn Gly Ile Lys Asp Ile Glu Glu Val Tyr  
 275 280 285  
 Leu Asn Gly Asp Leu Glu His Gly Ala Pro Asn Ile Leu Asn Val Ser  
 290 295 300  
 Phe Asn Tyr Val Glu Gly Glu Ser Leu Ile Met Ala Leu Lys Asp Leu  
 305 310 315 320  
 Ala Val Ser Ser Gly Ser Ala Cys Thr Ser Ala Ser Leu Glu Pro Ser  
 325 330 335  
 Tyr Val Leu Arg Ala Leu Gly Leu Asn Asp Glu Leu Ala His Ser Ser  
 340 345 350  
 Ile Arg Phe Ser Leu Gly Arg Phe Thr Thr Glu Glu Glu Ile Asp Tyr  
 355 360 365  
 Thr Ile Glu Leu Val Arg Lys Ser Ile Gly Arg Leu Arg Asp Leu Ser  
 370 375 380  
 Pro Leu Trp Glu Met Tyr Lys Gln Gly Val Asp Leu Asn Ser Ile Glu  
 385 390 395 400  
 Trp Ala His His

## (2) INFORMATION FOR SEQ ID No: 7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1221 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNS (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vi) ORIGINAL SOURCE:

(B) STRAIN: Escherichia coli

(vii) IMMEDIATE SOURCE:

(B) CLONE: bios3

## (ix) FEATURES:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1221

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 7:

ATG	ATT	TTT	TCC	GTC	GAC	AAA	GTG	CGG	GCC	GAC	TTT	CCG	GTG	CTT	TCG	48
Met	Ile	Phe	Ser	Val	Asp	Lys	Val	Arg	Ala	Asp	Phe	Pro	Val	Leu	Ser	
1				5					10					15		
CGT	GAG	GTA	AAC	GGT	TTG	CCG	CTG	GCT	TAT	CTC	GAC	AGC	GCC	GCC	AGT	96
Arg	Glu	Val	Asn	Gly	Leu	Pro	Leu	Ala	Tyr	Leu	Asp	Ser	Ala	Ala	Ser	
			20					25					30			
GCG	CAG	AAA	CCG	AGC	CAG	GTG	ATT	GAC	GCC	GAG	GCC	GAG	TTT	TAT	CGT	144
Ala	Gln	Lys	Pro	Ser	Gln	Val	Ile	Asp	Ala	Glu	Ala	Glu	Phe	Tyr	Arg	
		35					40					45				
CAT	GGC	TAC	GCG	GCG	GTG	CAT	CGT	GGT	ATT	CAT	ACC	TTA	AGC	GCC	CAG	192
His	Gly	Tyr	Ala	Ala	Val	His	Arg	Gly	Ile	His	Thr	Leu	Ser	Ala	Gln	
	50						55				60					
GCG	ACC	GAG	AAA	ATG	GAG	AAC	GTG	CGC	AAG	CGG	GCA	TCG	CTG	TTT	ATT	240
Ala	Thr	Glu	Lys	Met	Glu	Asn	Val	Arg	Lys	Arg	Ala	Ser	Leu	Phe	Ile	
65					70				75						80	
AAT	GCC	CGT	TCG	GCG	GAA	GAG	CTG	GTG	TTC	GTC	CGC	GGC	ACG	ACG	GAA	288
Asn	Ala	Arg	Ser	Ala	Glu	Glu	Leu	Val	Phe	Val	Arg	Gly	Thr	Thr	Glu	
				85					90					95		
GGG	ATC	AAT	CTG	GTC	GCC	AAT	AGC	TGG	GGC	AAC	AGC	AAC	GTG	CGG	GCG	336
Gly	Ile	Asn	Leu	Val	Ala	Asn	Ser	Trp	Gly	Asn	Ser	Asn	Val	Arg	Ala	
			100					105					110			
GGC	GAT	AAC	ATC	ATC	ATC	AGT	CAG	ATG	GAG	CAC	CAC	GCT	AAC	ATT	GTT	384
Gly	Asp	Asn	Ile	Ile	Ile	Ser	Gln	Met	Glu	His	His	Ala	Asn	Ile	Val	
		115					120					125				
CCC	TGG	CAG	ATG	CTT	TGC	GCA	CGC	GTT	GGC	GCA	GAG	CTG	CGT	GTG	ATC	432
Pro	Trp	Gln	Met	Leu	Cys	Ala	Arg	Val	Gly	Ala	Glu	Leu	Arg	Val	Ile	
	130						135					140				
CCG	CTC	AAT	CCC	GAT	GGT	ACG	TTG	CAA	CTG	GAG	ACG	CTG	CCT	ACG	CTG	480
Pro	Leu	Asn	Pro	Asp	Gly	Thr	Leu	Gln	Leu	Glu	Thr	Leu	Pro	Thr	Leu	
145					150					155					160	
TTT	GAT	GAG	AAA	ACT	CGC	CTG	CTG	GCA	ATT	ACT	CAT	GTC	TCC	AAC	GTG	528
Phe	Asp	Glu	Lys	Thr	Arg	Leu	Leu	Ala	Ile	Thr	His	Val	Ser	Asn	Val	
				165				170						175		

CTT	GGC	ACA	GAA	AAT	CCA	CTG	GCG	GAA	ATG	ATC	ACG	CTT	GCG	CAC	CAG	576
Leu	Gly	Thr	Glu	Asn	Pro	Leu	Ala	Glu	Met	Ile	Thr	Leu	Ala	His	Gln	
			180					185					190			
CAT	GGC	GCA	AAA	GTG	CTG	GTG	GAT	GGC	GCT	CAG	GCG	GTG	ATG	CAT	CAT	624
His	Gly	Ala	Lys	Val	Leu	Val	Asp	Gly	Ala	Gln	Ala	Val	Met	His	His	
		195					200					205				
CCG	GTG	GAT	GTT	CAG	GCG	CTG	GAT	TGC	GAC	TTT	TAC	GTG	TTC	TCC	GGG	672
Pro	Val	Asp	Val	Gln	Ala	Leu	Asp	Cys	Asp	Phe	Tyr	Val	Phe	Ser	Gly	
	210					215					220					
CAT	AAA	CTG	TAT	GGC	CCC	ACC	GGA	ATT	GGC	ATT	CTT	TAT	GTG	AAA	GAA	720
His	Lys	Leu	Tyr	Gly	Pro	Thr	Gly	Ile	Gly	Ile	Leu	Tyr	Val	Lys	Glu	
225					230					235					240	
GCC	TTG	TTG	CAG	GAG	ATG	CCG	CCG	TGG	GAA	GGG	GGC	GGT	TCT	ATG	ATC	768
Ala	Leu	Leu	Gln	Glu	Met	Pro	Pro	Trp	Glu	Gly	Gly	Gly	Ser	Met	Ile	
			245					250					255			
GCC	ACC	GTC	AGC	CTG	AGT	GAA	GGC	ACT	ACC	TGG	ACC	AAA	GCA	CCA	TGG	816
Ala	Thr	Val	Ser	Leu	Ser	Glu	Gly	Thr	Thr	Trp	Thr	Lys	Ala	Pro	Trp	
		260						265					270			
CGG	TTT	GAA	GCC	GGT	ACA	CCC	AAT	ACC	GGG	GGC	ATC	ATT	GGT	CTT	GGC	864
Arg	Phe	Glu	Ala	Gly	Thr	Pro	Asn	Thr	Gly	Gly	Ile	Ile	Gly	Leu	Gly	
	275						280					285				
GCG	GCG	CTG	GAG	TAT	GTT	TCG	GCG	CTG	GGG	CTT	AAT	AAC	ATA	GCC	GAG	912
Ala	Ala	Leu	Glu	Tyr	Val	Ser	Ala	Leu	Gly	Leu	Asn	Asn	Ile	Ala	Glu	
	290					295					300					
TAT	GAA	CAG	AAT	CTG	ATG	CAT	TAT	GCG	CTA	TCA	CAG	CTG	GAA	TCT	GTA	960
Tyr	Glu	Gln	Asn	Leu	Met	His	Tyr	Ala	Leu	Ser	Gln	Leu	Glu	Ser	Val	
305					310					315					320	
CCG	GAT	CTC	ACT	CTC	TAT	GGC	CCA	CAA	AAC	AGG	CTT	GGC	GTT	ATT	GCT	1008
Pro	Asp	Leu	Thr	Leu	Tyr	Gly	Pro	Gln	Asn	Arg	Leu	Gly	Val	Ile	Ala	
			325					330					335			
TTT	AAT	CTC	GGT	AAA	CAC	CAC	GCC	TAT	GAT	GTT	GGC	AGT	TTT	CTC	GAT	1056
Phe	Asn	Leu	Gly	Lys	His	His	Ala	Tyr	Asp	Val	Gly	Ser	Phe	Leu	Asp	
		340					345					350				
AAT	TAC	GGC	ATT	GCT	GTG	CGT	ACC	GGA	CAT	CAC	TGC	GCA	ATG	CCA	TTG	1104
Asn	Tyr	Gly	Ile	Ala	Val	Arg	Thr	Gly	His	His	Cys	Ala	Met	Pro	Leu	
		355					360					365				



ATG GCC TAT TAC AAC GTC CCT GCG ATG TGT CGG GCG TCG CTG GCC ATG 1152  
 Met Ala Tyr Tyr Asn Val Pro Ala Met Cys Arg Ala Ser Leu Ala Met  
 370 375 380

TAT AAC ACC CAT GAA GAA GTG GAT CGT CTG GTG ACC GGC CTG CAA CGT 1200  
 Tyr Asn Thr His Glu Glu Val Asp Arg Leu Val Thr Gly Leu Gln Arg  
 385 390 395 400

ATT CAC CGT TTG CTG GGA TAA 1221  
 Ile His Arg Leu Leu Gly  
 405

(2) INFORMATION FOR SEQ ID No: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 406 Amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 8:

Met Ile Phe Ser Val Asp Lys Val Arg Ala Asp Phe Pro Val Leu Ser  
 1 5 10 15

Arg Glu Val Asn Gly Leu Pro Leu Ala Tyr Leu Asp Ser Ala Ala Ser  
 20 25 30

Ala Gln Lys Pro Ser Gln Val Ile Asp Ala Glu Ala Glu Phe Tyr Arg  
 35 40 45

His Gly Tyr Ala Ala Val His Arg Gly Ile His Thr Leu Ser Ala Gln  
 50 55 60

Ala Thr Glu Lys Met Glu Asn Val Arg Lys Arg Ala Ser Leu Phe Ile  
 65 70 75 80

Asn Ala Arg Ser Ala Glu Glu Leu Val Phe Val Arg Gly Thr Thr Glu  
 85 90 95

Gly Ile Asn Leu Val Ala Asn Ser Trp Gly Asn Ser Asn Val Arg Ala  
 100 105 110

Gly Asp Asn Ile Ile Ile Ser Gln Met Glu His His Ala Asn Ile Val  
 115 120 125

Pro Trp Gln Met Leu Cys Ala Arg Val Gly Ala Glu Leu Arg Val Ile  
 130 135 140

Pro Leu Asn Pro Asp Gly Thr Leu Gln Leu Glu Thr Leu Pro Thr Leu  
145 150 155 160

Phe Asp Glu Lys Thr Arg Leu Leu Ala Ile Thr His Val Ser Asn Val  
165 170 175

Leu Gly Thr Glu Asn Pro Leu Ala Glu Met Ile Thr Leu Ala His Gln  
180 185 190

His Gly Ala Lys Val Leu Val Asp Gly Ala Gln Ala Val Met His His  
195 200 205

Pro Val Asp Val Gln Ala Leu Asp Cys Asp Phe Tyr Val Phe Ser Gly  
210 215 220

His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Ile Leu Tyr Val Lys Glu  
225 230 235 240

Ala Leu Leu Gln Glu Met Pro Pro Trp Glu Gly Gly Gly Ser Met Ile  
245 250 255

Ala Thr Val Ser Leu Ser Glu Gly Thr Thr Trp Thr Lys Ala Pro Trp  
260 265 270

Arg Phe Glu Ala Gly Thr Pro Asn Thr Gly Gly Ile Ile Gly Leu Gly  
275 280 285

Ala Ala Leu Glu Tyr Val Ser Ala Leu Gly Leu Asn Asn Ile Ala Glu  
290 295 300

Tyr Glu Gln Asn Leu Met His Tyr Ala Leu Ser Gln Leu Glu Ser Val  
305 310 315 320

Pro Asp Leu Thr Leu Tyr Gly Pro Gln Asn Arg Leu Gly Val Ile Ala  
325 330 335

Phe Asn Leu Gly Lys His His Ala Tyr Asp Val Gly Ser Phe Leu Asp  
340 345 350

Asn Tyr Gly Ile Ala Val Arg Thr Gly His His Cys Ala Met Pro Leu  
355 360 365

Met Ala Tyr Tyr Asn Val Pro Ala Met Cys Arg Ala Ser Leu Ala Met  
370 375 380

Tyr Asn Thr His Glu Glu Val Asp Arg Leu Val Thr Gly Leu Gln Arg  
385 390 395 400

Ile His Arg Leu Leu Gly  
405

## (2) INFORMATION FOR SEQ ID No: 9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3720 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: pHS1 metK

## (ix) FEATURES:

- (A) NAME/KEY: CDS
- (B) LOCATION: 530..1684

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 9:

GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTACAC AGGGCCCTCG GACACCGAGG 60  
 AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC 120  
 GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA 180  
 AGGTAGTTAT CCTTAAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA 240  
 AGCTCATTTT TGAAGAGGAC TTGTTGCGGA AACGACGAGA ACAGTTGAAA CACAAACTTG 300  
 AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCTT TCTAACAGAA 360  
 ATGTCCTGAG CAATCACCTA TGAAGTGTG ACTCGAGATA GCATTTTTAT CCATAAGATT 420  
 AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC 480  
 GGATAACAAT TTCACACACG CTAGCGGTAC CAAAGAGGAG AAATTAAC ATG GCA 535  
 Met Ala

1

AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT GAC AAA 583  
 Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro Asp Lys  
 5 10 15

ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC GAA CAG 631  
 Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu Glu Gln  
 20 25 30

GAT	CCG	AAA	GCA	CGC	GTT	GCT	TGC	GAA	ACC	TAC	GTA	AAA	ACC	GGC	ATG	679
Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr	Gly	Met	
35					40					45					50	
GTT	TTA	GTT	GGC	GGC	GAA	ATC	ACC	ACC	AGC	GCC	TGG	GTA	GAC	ATC	GAA	727
Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp	Ile	Glu	
			55						60					65		
GAG	ATC	ACC	CGT	AAC	ACC	GTT	CGC	GAA	ATT	GGC	TAT	GTG	CAT	TCC	GAC	775
Glu	Ile	Thr	Arg	Asn	Thr	Val	Arg	Glu	Ile	Gly	Tyr	Val	His	Ser	Asp	
			70					75					80			
ATG	GGC	TTT	GAC	GCT	AAC	TCC	TGT	GCG	GTT	CTG	AGC	GCT	ATC	GGC	AAA	823
Met	Gly	Phe	Asp	Ala	Asn	Ser	Cys	Ala	Val	Leu	Ser	Ala	Ile	Gly	Lys	
	85						90					95				
CAG	TCT	CCT	GAC	ATC	AAC	CAG	GGC	GTT	GAC	CGT	GCC	GAT	CCG	CTG	GAA	871
Gln	Ser	Pro	Asp	Ile	Asn	Gln	Gly	Val	Asp	Arg	Ala	Asp	Pro	Leu	Glu	
	100					105					110					
CAG	GGC	GCG	GGT	GAC	CAG	GGT	CTG	ATG	TTT	GGC	TAC	GCA	ACT	AAT	GAA	919
Gln	Gly	Ala	Gly	Asp	Gln	Gly	Leu	Met	Phe	Gly	Tyr	Ala	Thr	Asn	Glu	
115					120					125					130	
ACC	GAC	GTG	CTG	ATG	CCA	GCA	CCT	ATC	ACC	TAT	GCA	CAC	CGT	CTG	GTA	967
Thr	Asp	Val	Leu	Met	Pro	Ala	Pro	Ile	Thr	Tyr	Ala	His	Arg	Leu	Val	
				135					140					145		
CAG	CGT	CAG	GCT	GAA	GTG	CGT	AAA	AAC	GGC	ACT	CTG	CCG	TGG	CTG	CGC	1015
Gln	Arg	Gln	Ala	Glu	Val	Arg	Lys	Asn	Gly	Thr	Leu	Pro	Trp	Leu	Arg	
			150					155					160			
CCG	GAC	GCG	AAA	AGC	CAG	GTG	ACT	TTT	CAG	TAT	GAC	GAC	GGC	AAA	ATC	1063
Pro	Asp	Ala	Lys	Ser	Gln	Val	Thr	Phe	Gln	Tyr	Asp	Asp	Gly	Lys	Ile	
	165						170					175				
GTT	GGT	ATC	GAT	GCT	GTC	GTG	CTT	TCC	ACT	CAG	CAC	TCT	GAA	GAG	ATC	1111
Val	Gly	Ile	Asp	Ala	Val	Val	Leu	Ser	Thr	Gln	His	Ser	Glu	Glu	Ile	
	180						185					190				
GAC	CAG	AAA	TCG	CTG	CAA	GAA	GCG	GTA	ATG	GAA	GAG	ATC	ATC	AAG	CCA	1159
Asp	Gln	Lys	Ser	Leu	Gln	Glu	Ala	Val	Met	Glu	Glu	Ile	Ile	Lys	Pro	
195					200					205					210	
ATT	CTG	CCC	GCT	GAA	TGG	CTG	ACT	TCT	GCC	ACC	AAA	TTC	TTC	ATC	AAC	1207
Ile	Leu	Pro	Ala	Glu	Trp	Leu	Thr	Ser	Ala	Thr	Lys	Phe	Phe	Ile	Asn	
			215						220					225		

CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC GGT CTG	1255
Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys Gly Leu	
230 235 240	
ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG CGT CAC	1303
Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala Arg His	
245 250 255	
GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC CGT TCC	1351
Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp Arg Ser	
260 265 270	
GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT GCT GGC	1399
Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala Ala Gly	
275 280 285 290	
CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC GTG GCT	1447
Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly Val Ala	
295 300 305	
GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA GTG CCT	1495
Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys Val Pro	
310 315 320	
TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG CGC CCA	1543
Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu Arg Pro	
325 330 335	
TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC AAA GAA	1591
Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr Lys Glu	
340 345 350	
ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG GAA AAA	1639
Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp Glu Lys	
355 360 365 370	
ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG TAATCGGTAC	1691
Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys	
375 380 385	
CGCTTGATAT CGAATTCCTG CAGCCCGGGG GATCCCATGG TACGCGTGCT AGAGGCATCA	1751
AATAAAACGA AAGGCTCAGT CGAAAGACTG GGCCTTTCGT TTTATCTGTT GTTTGTCGGT	1811
GAACGCTCTC CTGAGTAGGA CAAATCCGCC GCCCTAGACC TAGGGGATAT ATTCCGCTTC	1871
CTCGCTCACT GACTCGCTAC GCTCGGTCGT TCGACTGCGG CGAGCGGAAA TGGCTTACGA	1931
ACGGGGCGGA GATTCCTGG AAGATGCCAG GAAGATACTT AACAGGGAAG TGAGAGGGCC	1991

GCGGCAAAGC	CGTTTTTCCA	TAGGCTCCGC	CCCCCTGACA	AGCATCACGA	AATCTGACGC	2051
TCAAATCAGT	GGTGGCGAAA	CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGC	2111
GGCTCCCTCG	TGCGCTCTCC	TGTTCCCTGCC	TTTCGGTTTA	CCGGTGTCAT	TCCGCTGTTA	2171
TGGCCGCGTT	TGTCTCATTC	CACGCCTGAC	ACTCAGTTCC	GGGTAGGCAG	TTCGCTCCAA	2231
GCTGGACTGT	ATGCACGAAC	CCCCCGTTCA	GTCCGACCGC	TGCGCCTTAT	CCGGTAACTA	2291
TCGTCTTGAG	TCCAACCCGG	AAAGACATGC	AAAAGCACCA	CTGGCAGCAG	CCACTGGTAA	2351
TTGATTTAGA	GGAGTTAGTC	TTGAAGTCAT	GCGCCGGTTA	AGGCTAAACT	GAAAGGACAA	2411
GTTTTGGTGA	CTGCGCTCCT	CCAAGCCAGT	TACCTCGGTT	CAAAGAGTTG	GTAGCTCAGA	2471
GAACCTTCGA	AAAACCGCCC	TGCAAGGCGG	TTTTTTCGTT	TTCAGAGCAA	GAGATTACGC	2531
GCAGACCAAA	ACGATCTCAA	GAAGATCATC	TTATTAATCA	GATAAAATAT	TTCTAGATTT	2591
CAGTGCAATT	TATCTCTTCA	AATGTAGCAC	CTGAAGTCAG	CCCCATACGA	TATAAGTTGT	2651
TACTAGTGCT	TGGATTCTCA	CCAATAAAAA	ACGCCCCGGC	GCAACCGAGC	GTTCTGAACA	2711
AATCCAGATG	GAGTTCTGAG	GTCATTACTG	GATCTATCAA	CAGGAGTCCA	AGCGAGCTCT	2771
CGAACCCAG	AGTCCCGCTC	AGAAGAACTC	GTCAAGAAGG	CGATAGAAGG	CGATGCGCTG	2831
CGAATCGGGA	GCGGCGATAC	CGTAAAGCAC	GAGGAAGCGG	TCAGCCCATT	CGCCGCCAAG	2891
CTCTTCAGCA	ATATCACGGG	TAGCCAACGC	TATGTCCTGA	TAGCGGTCCG	CCACACCCAG	2951
CCGGCCACAG	TCGATGAATC	CAGAAAAGCG	GCCATTTTCC	ACCATGATAT	TCGGCAAGCA	3011
GGCATCGCCA	TGGGTCACGA	CGAGATCCTC	GCCGTCGGGC	ATGCGCGCCT	TGAGCCTGGC	3071
GAACAGTTTC	GCTGGCGCGA	GCCCCTGATG	CTCTTCGTCC	AGATCATCCT	GATCGACAAG	3131
ACCGGCTTCC	ATCCGAGTAC	GTGCTCGCTC	GATGCGATGT	TTCGCTTGGT	GGTCGAATGG	3191
GCAGGTAGCC	GGATCAAGCG	TATGCAGCCG	CCGCATTGCA	TCAGCCATGA	TGGATACTTT	3251
CTCGGCAGGA	GCAAGGTGAG	ATGACAGGAG	ATCCTGCCCC	GGCACTTCGC	CCAATAGCAG	3311
CCAGTCCCTT	CCCGCTTCAG	TGACAACGTC	GAGCACAGCT	GCGCAAGGAA	CGCCCGTCGT	3371
GGCCAGCCAC	GATAGCCGCG	CTGCCTCGTC	CTGCAGTTCA	TTCAGGGCAC	CGGACAGGTC	3431
GGTCTTGACA	AAAAGAACCG	GGCGCCCCTG	CGCTGACAGC	CGGAACACGG	CGGCATCAGA	3491

GCAGCCGATT GTCTGTTGTG CCCAGTCATA GCCGAATAGC CTCTCCACCC AAGCGGCCGG 3551  
 AGAACCTGCG TGCAATCCAT CTTGTTCAAT CATGCGAAAC GATCCTCATC CTGTCTCTTG 3611  
 ATCAGATCTT GATCCCCTGC GCCATCAGAT CCTTGGCGGC AAGAAAGCCA TCCAGTTTAC 3671  
 TTTGCAGGGC TTCCCAACCT TACCAGAGGG CGCCCCAGCT GGCAATTCC 3720

## (2) INFORMATION FOR SEQ ID No: 10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 384 Amino acids  
 (B) TYPE: Amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 10:

Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro  
 1 5 10 15  
 Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu  
 20 25 30  
 Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr  
 35 40 45  
 Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp  
 50 55 60  
 Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His  
 65 70 75 80  
 Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile  
 85 90 95  
 Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro  
 100 105 110  
 Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr  
 115 120 125  
 Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg  
 130 135 140  
 Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp  
 145 150 155 160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly  
 165 170 175  
 Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu  
 180 185 190  
 Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile  
 195 200 205  
 Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe  
 210 215 220  
 Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys  
 225 230 235 240  
 Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala  
 245 250 255  
 Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp  
 260 265 270  
 Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala  
 275 280 285  
 Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly  
 290 295 300  
 Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys  
 305 310 315 320  
 Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu  
 325 330 335  
 Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr  
 340 345 350  
 Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp  
 355 360 365  
 Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys  
 370 375 380

(2) INFORMATION FOR SEQ ID No: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3794 Base pairs
- (B) TYPE: Nucleic acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: circular



(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: pHS1 bioS1

(ix) FEATURES:

(A) NAME/KEY: CDS

(B) LOCATION: 601..1806

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 11:

GACGTCTGTG	TGGAATTGTG	AGCGGATAAC	AATTTACAC	AGGGCCCTCG	GACACCGAGG	60
AGAATGTCAA	GAGGCGAACA	CACAACGTCT	TGGAGCGCCA	GAGGAGGAAC	GAGCTAAAAC	120
GGAGCTTTTT	TGCCCTGCGT	GACCAGATCC	CGGAGTTGGA	AAACAATGAA	AAGGCCCCCA	180
AGGTAGTTAT	CCTTAAAAAA	GCCACAGCAT	ACATCCTGTC	CGTCCAAGCA	GAGGAGCAAA	240
AGCTCATTTT	TGAAGAGGAC	TTGTTGCGGA	AACGACGAGA	ACAGTTGAAA	CACAAACTTG	300
AACAGCTACG	GAATCTTGT	GCGTAAGGAA	AAGTAAGGAA	AACGATTCCCT	TCTAACAGAA	360
ATGTCCTGAG	CAATCACCTA	TGAACTGTG	ACTCGAGATA	GCATTTTAT	CCATAAGATT	420
AGCCGATCCT	AAGGTTTACA	ATTGTGAGCG	CTCACAATTA	TGATAGATTC	AATTGTGAGC	480
GGATAACAAT	TTCACACACG	CTAGCGGTAC	CGGGCCCCCC	CTCGAGGTCG	ACGGTATCGA	540
TAAGCTTGAT	ATCGAATTCC	TGCAGCCCGG	GGGATCCCAT	GGTACGCGTC	GAGGAGTACC	600
ATG AAC GTT TTT AAT CCC GCG CAG TTT CGC GCC CAG TTT CCC GCA CTA	648					
Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu						
1 5 10 15						
CAG GAT GCG GGC GTC TAT CTC GAC AGC GCC GCG ACC GCG CTT AAA CCT	696					
Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro						
20 25 30						
GAA GCC GTG GTT GAA GCC ACC CAA CAG TTT TAC AGT CTG AGC GCC GGA	744					
Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly						
35 40 45						
AAC GTC CAT CGC AGC CAG TTT GCC GAA GCC CAA CGC CTG ACC GCG CGT	792					
Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg						
50 55 60						

TAT	GAA	GCT	GCA	CGA	GAG	AAA	GTG	GCG	CAA	TTA	CTG	AAT	GCA	CCG	GAT	840
Tyr	Glu	Ala	Ala	Arg	Glu	Lys	Val	Ala	Gln	Leu	Leu	Asn	Ala	Pro	Asp	
65					70				75						80	
GAT	AAA	ACT	ATC	GTC	TGG	ACG	CGC	GGC	ACC	ACT	GAA	TCC	ATC	AAC	ATG	888
Asp	Lys	Thr	Ile	Val	Trp	Thr	Arg	Gly	Thr	Thr	Glu	Ser	Ile	Asn	Met	
				85				90						95		
GTG	GCA	CAA	TGC	TAT	GCG	CGT	CCG	CGT	CTG	CAA	CCG	GGC	GAT	GAG	ATT	936
Val	Ala	Gln	Cys	Tyr	Ala	Arg	Pro	Arg	Leu	Gln	Pro	Gly	Asp	Glu	Ile	
			100					105					110			
ATT	GTC	AGC	GTG	GCA	GAA	CAC	CAC	GCC	AAC	CTC	GTC	CCC	TGG	CTG	ATG	984
Ile	Val	Ser	Val	Ala	Glu	His	His	Ala	Asn	Leu	Val	Pro	Trp	Leu	Met	
		115					120					125				
GTC	GCC	CAA	CAA	ACT	GGA	GCC	AAA	GTG	GTG	AAA	TTG	CCG	CTT	AAT	GCG	1032
Val	Ala	Gln	Gln	Thr	Gly	Ala	Lys	Val	Val	Lys	Leu	Pro	Leu	Asn	Ala	
	130					135					140					
CAG	CGA	CTG	CCG	GAT	GTC	GAT	TTG	TTG	CCA	GAA	CTG	ATT	ACT	CCC	CGT	1080
Gln	Arg	Leu	Pro	Asp	Val	Asp	Leu	Leu	Pro	Glu	Leu	Ile	Thr	Pro	Arg	
145					150					155					160	
AGT	CGG	ATT	CTG	GCG	TTG	GGT	CAG	ATG	TCG	AAC	GTT	ACT	GGC	GGT	TGC	1128
Ser	Arg	Ile	Leu	Ala	Leu	Gly	Gln	Met	Ser	Asn	Val	Thr	Gly	Gly	Cys	
				165				170						175		
CCG	GAT	CTG	GCG	CGA	GCG	ATT	ACC	TTT	GCT	CAT	TCA	GCC	GGG	ATG	GTG	1176
Pro	Asp	Leu	Ala	Arg	Ala	Ile	Thr	Phe	Ala	His	Ser	Ala	Gly	Met	Val	
			180					185					190			
GTG	ATG	GTT	GAT	GGT	GCT	CAG	GGG	GCA	GTG	CAT	TTC	CCC	GCG	GAT	GTT	1224
Val	Met	Val	Asp	Gly	Ala	Gln	Gly	Ala	Val	His	Phe	Pro	Ala	Asp	Val	
		195					200					205				
CAG	CAA	CTG	GAT	ATT	GAT	TTC	TAT	GCT	TTT	TCA	GGT	CAC	AAA	CTG	TAT	1272
Gln	Gln	Leu	Asp	Ile	Asp	Phe	Tyr	Ala	Phe	Ser	Gly	His	Lys	Leu	Tyr	
	210					215					220					
GGC	CCG	ACA	GGT	ATC	GGC	GTG	CTG	TAT	GGT	AAA	TCA	GAA	CTG	CTG	GAG	1320
Gly	Pro	Thr	Gly	Ile	Gly	Val	Leu	Tyr	Gly	Lys	Ser	Glu	Leu	Leu	Glu	
225					230					235					240	
GCG	ATG	TCG	CCC	TGG	CTG	GGC	GGC	GGC	AAA	ATG	GTT	CAC	GAA	GTG	AGT	1368
Ala	Met	Ser	Pro	Trp	Leu	Gly	Gly	Gly	Lys	Met	Val	His	Glu	Val	Ser	
				245				250						255		

TTT GAC GGC TTC ACG ACT CAA TCT GCG CCG TGG AAA CTG GAA GCT GGA	1416
Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly	
260 265 270	
ACG CCA AAT GTC GCT GGT GTC ATA GGA TTA AGC GCG GCG CTG GAA TGG	1464
Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp	
275 280 285	
CTG GCA GAT TAC GAT ATC AAC CAG GCC GAA AGC TGG AGC CGT AGC TTA	1512
Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu	
290 295 300	
GCA ACG CTG GCG GAA GAT GCG CTG GCG AAA CGT CCC GGC TTT CGT TCA	1560
Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser	
305 310 315 320	
TTC CGC TGC CAG GAT TCC AGC CTG CTG GCC TTT GAT TTT GCT GGC GTT	1608
Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val	
325 330 335	
CAT CAT AGC GAT ATG GTG ACG CTG CTG GCG GAG TAC GGT ATT GCC CTG	1656
His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu	
340 345 350	
CGG GCC GGG CAG CAT TGC GCT CAG CCG CTA CTG GCA GAA TTA GGC GTA	1704
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val	
355 360 365	
ACC GGC ACA CTG CGC GCC TCT TTT GCG CCA TAT AAT ACA AAG AGT GAT	1752
Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp	
370 375 380	
GTG GAT GCG CTG GTG AAT GCC GTT GAC CGC GCG CTG GAA TTA TTG GTG	1800
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val	
385 390 395 400	
GAT TAAACGCGTG CTAGAGGCAT CAAATAAAAC GAAAGGCTCA GTCGAAAGAC	1853
Asp	
TGGGCCTTTC GTTTTATCTG TTGTTTGTCTG GTGAACGCTC TCCTGAGTAG GACAAATCCG	1913
CCGCCCTAGA CCTAGGGGAT ATATTCCGCT TCCTCGCTCA CTGACTCGCT ACGCTCGGTC	1973
GTTCGACTGC GGCGAGCGGA AATGGCTTAC GAACGGGGCG GAGATTTTCT GGAAGATGCC	2033
AGGAAGATAC TTAACAGGGA AGTGAGAGGG CCGCGGCAAA GCCGTTTTTC CATAGGCTCC	2093
GCCCCCTGA CAAGCATCAC GAAATCTGAC GCTCAAATCA GTGGTGGCGA AACCCGACAG	2153

GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	GCGGCTCCCT	CGTGCGCTCT	CCTGTTCCCTG	2213
CCTTTTCGGTT	TACCGGTGTC	ATTCCGCTGT	TATGGCCGCG	TTTGTCTCAT	TCCACGCCTG	2273
ACACTCAGTT	CCGGGTAGGC	AGTTCGCTCC	AAGCTGGACT	GTATGCACGA	ACCCCCCGTT	2333
CAGTCCGACC	GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	AGTCCAACCC	GGAAAGACAT	2393
GCAAAAGCAC	CACTGGCAGC	AGCCACTGGT	AATTGATTTA	GAGGAGTTAG	TCTTGAAGTC	2453
ATGCGCCGGT	TAAGGCTAAA	CTGAAAGGAC	AAGTTTTGGT	GACTGCGCTC	CTCCAAGCCA	2513
GTTACCTCGG	TTCAAAGAGT	TGGTAGCTCA	GAGAACCTTC	GAAAAACCGC	CCTGCAAGGC	2573
GGTTTTTTTCG	TTTTTCAGAGC	AAGAGATTAC	GCGCAGACCA	AAACGATCTC	AAGAAGATCA	2633
TCTTATTAAT	CAGATAAAAT	ATTTCTAGAT	TTCAGTGCAA	TTTATCTCTT	CAAATGTAGC	2693
ACCTGAAGTC	AGCCCCATAC	GATATAAGTT	GTTACTAGTG	CTTGGATTCT	CACCAATAAA	2753
AAACGCCCCG	CGGCAACCGA	GCGTTCTGAA	CAAATCCAGA	TGGAGTTCTG	AGGTCATTAC	2813
TGGATCTATC	AACAGGAGTC	CAAGCGAGCT	CTCGAACCCC	AGAGTCCCGC	TCAGAAGAAC	2873
TCGTCAAGAA	GGCGATAGAA	GGCGATGCGC	TGCGAATCGG	GAGCGGCGAT	ACCGTAAAGC	2933
ACGAGGAAGC	GGTCAGCCCA	TTCGCCGCCA	AGCTCTTCAG	CAATATCACG	GGTAGCCAAC	2993
GCTATGTCCT	GATAGCGGTC	CGCCACACCC	AGCCGGCCAC	AGTCGATGAA	TCCAGAAAAG	3053
CGGCCATTTT	CCACCATGAT	ATTCGGCAAG	CAGGCATCGC	CATGGGTCAC	GACGAGATCC	3113
TCGCCGTCGG	GCATGCGCGC	CTTGAGCCTG	GCGAACAGTT	CGGCTGGCGC	GAGCCCCTGA	3173
TGCTCTTCGT	CCAGATCATC	CTGATCGACA	AGACCGGCTT	CCATCCGAGT	ACGTGCTCGC	3233
TCGATGCGAT	GTTTCGCTTG	GTGGTCGAAT	GGGCAGGTAG	CCGGATCAAG	CGTATGCAGC	3293
CGCCGCATTG	CATCAGCCAT	GATGGATACT	TTCTCGGCAG	GAGCAAGGTG	AGATGACAGG	3353
AGATCCTGCC	CCGGCACTTC	GCCCAATAGC	AGCCAGTCCC	TTCCCGCTTC	AGTGACAACG	3413
TCGAGCACAG	CTGCGCAAGG	AACGCCCGTC	GTGGCCAGCC	ACGATAGCCG	CGCTGCCTCG	3473
TCCTGCAGTT	CATTCAGGGC	ACCGGACAGG	TCGGTCTTGA	CAAAAAGAAC	CGGGCGCCCC	3533
TGCGCTGACA	GCCGGAACAC	GGCGGCATCA	GAGCAGCCGA	TTGTCTGTTG	TGCCCAGTCA	3593
TAGCCGAATA	GCCTCTCCAC	CCAAGCGGCC	GGAGAACCTG	CGTGCAATCC	ATCTTGTTCA	3653

ATCATGCGAA ACGATCCTCA TCCTGTCTCT TGATCAGATC TTGATCCCCT GCGCCATCAG 3713  
 ATCCTTGGCG GCAAGAAAGC CATCCAGTTT ACTTTGCAGG GCTTCCCAAC CTTACCAGAG 3773  
 GGCGCCCCAG CTGGCAATTC C 3794

## (2) INFORMATION FOR SEQ ID No: 12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 401 Amino acids  
 (B) TYPE: Amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 12:

Met	Asn	Val	Phe	Asn	Pro	Ala	Gln	Phe	Arg	Ala	Gln	Phe	Pro	Ala	Leu	1	5	10	15
Gln	Asp	Ala	Gly	Val	Tyr	Leu	Asp	Ser	Ala	Ala	Thr	Ala	Leu	Lys	Pro	20	25	30	
Glu	Ala	Val	Val	Glu	Ala	Thr	Gln	Gln	Phe	Tyr	Ser	Leu	Ser	Ala	Gly	35	40	45	
Asn	Val	His	Arg	Ser	Gln	Phe	Ala	Glu	Ala	Gln	Arg	Leu	Thr	Ala	Arg	50	55	60	
Tyr	Glu	Ala	Ala	Arg	Glu	Lys	Val	Ala	Gln	Leu	Leu	Asn	Ala	Pro	Asp	65	70	75	80
Asp	Lys	Thr	Ile	Val	Trp	Thr	Arg	Gly	Thr	Thr	Glu	Ser	Ile	Asn	Met	85	90	95	
Val	Ala	Gln	Cys	Tyr	Ala	Arg	Pro	Arg	Leu	Gln	Pro	Gly	Asp	Glu	Ile	100	105	110	
Ile	Val	Ser	Val	Ala	Glu	His	His	Ala	Asn	Leu	Val	Pro	Trp	Leu	Met	115	120	125	
Val	Ala	Gln	Gln	Thr	Gly	Ala	Lys	Val	Val	Lys	Leu	Pro	Leu	Asn	Ala	130	135	140	
Gln	Arg	Leu	Pro	Asp	Val	Asp	Leu	Leu	Pro	Glu	Leu	Ile	Thr	Pro	Arg	145	150	155	160
Ser	Arg	Ile	Leu	Ala	Leu	Gly	Gln	Met	Ser	Asn	Val	Thr	Gly	Gly	Cys	165	170	175	

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val  
 180 185 190  
 Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val  
 195 200 205  
 Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr  
 210 215 220  
 Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu  
 225 230 235 240  
 Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser  
 245 250 255  
 Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly  
 260 265 270  
 Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp  
 275 280 285  
 Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu  
 290 295 300  
 Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser  
 305 310 315 320  
 Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val  
 325 330 335  
 His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu  
 340 345 350  
 Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val  
 355 360 365  
 Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp  
 370 375 380  
 Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val  
 385 390 395 400  
 Asp

## (2) INFORMATION FOR SEQ ID No: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4975 Base pairs
- (B) TYPE: Nucleic acid

(C) STRANDEDNESS: Single  
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTISENSE: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: pHS1 metK bioS1

(ix) FEATURES:

(A) NAME/KEY: CDS

(B) LOCATION: 1782..2987

(ix) FEATURES:

(A) NAME/KEY: CDS

(B) LOCATION: 530..1684

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 13:

GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTACAC AGGGCCCTCG GACACCGAGG	60
AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC	120
GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA	180
AGGTAGTTAT CCTTAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA	240
AGCTCATTTT TGAAGAGGAC TTGTTGCGGA AACGACGAGA ACAGTTGAAA CACAACTTG	300
AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCCT TCTAACAGAA	360
ATGTCCTGAG CAATCACCTA TGAAGTGTG ACTCGAGATA GCATTTTAT CCATAAGATT	420
AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC	480
GGATAACAAT TTCACACACG CTAGCGGTAC CAAAGAGGAG AAATTAACAT ATG GCA	535
	Met Ala
	1
AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT GAC AAA	583
Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro Asp Lys	
5 10 15	
ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC GAA CAG	631
Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu Glu Gln	
20 25 30	

GAT	CCG	AAA	GCA	CGC	GTT	GCT	TGC	GAA	ACC	TAC	GTA	AAA	ACC	GGC	ATG	679
Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr	Gly	Met	
35					40					45					50	
GTT	TTA	GTT	GGC	GGC	GAA	ATC	ACC	ACC	AGC	GCC	TGG	GTA	GAC	ATC	GAA	727
Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp	Ile	Glu	
			55						60					65		
GAG	ATC	ACC	CGT	AAC	ACC	GTT	CGC	GAA	ATT	GGC	TAT	GTG	CAT	TCC	GAC	775
Glu	Ile	Thr	Arg	Asn	Thr	Val	Arg	Glu	Ile	Gly	Tyr	Val	His	Ser	Asp	
			70					75					80			
ATG	GGC	TTT	GAC	GCT	AAC	TCC	TGT	GCG	GTT	CTG	AGC	GCT	ATC	GGC	AAA	823
Met	Gly	Phe	Asp	Ala	Asn	Ser	Cys	Ala	Val	Leu	Ser	Ala	Ile	Gly	Lys	
	85						90					95				
CAG	TCT	CCT	GAC	ATC	AAC	CAG	GGC	GTT	GAC	CGT	GCC	GAT	CCG	CTG	GAA	871
Gln	Ser	Pro	Asp	Ile	Asn	Gln	Gly	Val	Asp	Arg	Ala	Asp	Pro	Leu	Glu	
100					105					110						
CAG	GGC	GCG	GGT	GAC	CAG	GGT	CTG	ATG	TTT	GGC	TAC	GCA	ACT	AAT	GAA	919
Gln	Gly	Ala	Gly	Asp	Gln	Gly	Leu	Met	Phe	Gly	Tyr	Ala	Thr	Asn	Glu	
115					120					125					130	
ACC	GAC	GTG	CTG	ATG	CCA	GCA	CCT	ATC	ACC	TAT	GCA	CAC	CGT	CTG	GTA	967
Thr	Asp	Val	Leu	Met	Pro	Ala	Pro	Ile	Thr	Tyr	Ala	His	Arg	Leu	Val	
				135					140					145		
CAG	CGT	CAG	GCT	GAA	GTG	CGT	AAA	AAC	GGC	ACT	CTG	CCG	TGG	CTG	CGC	1015
Gln	Arg	Gln	Ala	Glu	Val	Arg	Lys	Asn	Gly	Thr	Leu	Pro	Trp	Leu	Arg	
			150				155						160			
CCG	GAC	GCG	AAA	AGC	CAG	GTG	ACT	TTT	CAG	TAT	GAC	GAC	GGC	AAA	ATC	1063
Pro	Asp	Ala	Lys	Ser	Gln	Val	Thr	Phe	Gln	Tyr	Asp	Asp	Gly	Lys	Ile	
	165						170					175				
GTT	GGT	ATC	GAT	GCT	GTC	GTG	CTT	TCC	ACT	CAG	CAC	TCT	GAA	GAG	ATC	1111
Val	Gly	Ile	Asp	Ala	Val	Val	Leu	Ser	Thr	Gln	His	Ser	Glu	Glu	Ile	
	180						185					190				
GAC	CAG	AAA	TCG	CTG	CAA	GAA	GCG	GTA	ATG	GAA	GAG	ATC	ATC	AAG	CCA	1159
Asp	Gln	Lys	Ser	Leu	Gln	Glu	Ala	Val	Met	Glu	Glu	Ile	Ile	Lys	Pro	
195					200					205					210	
ATT	CTG	CCC	GCT	GAA	TGG	CTG	ACT	TCT	GCC	ACC	AAA	TTC	TTC	ATC	AAC	1207
Ile	Leu	Pro	Ala	Glu	Trp	Leu	Thr	Ser	Ala	Thr	Lys	Phe	Phe	Ile	Asn	
			215						220					225		



CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC GGT CTG	1255
Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys Gly Leu	
230 235 240	
ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG CGT CAC	1303
Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala Arg His	
245 250 255	
GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC CGT TCC	1351
Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp Arg Ser	
260 265 270	
GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT GCT GGC	1399
Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala Ala Gly	
275 280 285 290	
CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC GTG GCT	1447
Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly Val Ala	
295 300 305	
GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA GTG CCT	1495
Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys Val Pro	
310 315 320	
TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG CGC CCA	1543
Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu Arg Pro	
325 330 335	
TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC AAA GAA	1591
Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr Lys Glu	
340 345 350	
ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG GAA AAA	1639
Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp Glu Lys	
355 360 365 370	
ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG TAATCGGTAC	1691
Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys	
375 380 385	
CGGGCCCCC CTCGAGGTCG ACGGTATCGA TAAGCTTGAT ATCGAATTCC TGCAGCCCCG	1751
GGGATCCCAT GGTACGCGTC GAGGAGTACC ATG AAC GTT TTT AAT CCC GCG CAG	1805
Met Asn Val Phe Asn Pro Ala Gln	
1 5	
TTT CGC GCC CAG TTT CCC GCA CTA CAG GAT GCG GGC GTC TAT CTC GAC	1853
Phe Arg Ala Gln Phe Pro Ala Leu Gln Asp Ala Gly Val Tyr Leu Asp	
10 15 20	

AGC	GCC	GCG	ACC	GCG	CTT	AAA	CCT	GAA	GCC	GTG	GTT	GAA	GCC	ACC	CAA	1901
Ser	Ala	Ala	Thr	Ala	Leu	Lys	Pro	Glu	Ala	Val	Val	Glu	Ala	Thr	Gln	
25					30					35					40	
CAG	TTT	TAC	AGT	CTG	AGC	GCC	GGA	AAC	GTC	CAT	CGC	AGC	CAG	TTT	GCC	1949
Gln	Phe	Tyr	Ser	Leu	Ser	Ala	Gly	Asn	Val	His	Arg	Ser	Gln	Phe	Ala	
				45					50					55		
GAA	GCC	CAA	CGC	CTG	ACC	GCG	CGT	TAT	GAA	GCT	GCA	CGA	GAG	AAA	GTG	1997
Glu	Ala	Gln	Arg	Leu	Thr	Ala	Arg	Tyr	Glu	Ala	Ala	Arg	Glu	Lys	Val	
			60					65					70			
GCG	CAA	TTA	CTG	AAT	GCA	CCG	GAT	GAT	AAA	ACT	ATC	GTC	TGG	ACG	CGC	2045
Ala	Gln	Leu	Leu	Asn	Ala	Pro	Asp	Asp	Lys	Thr	Ile	Val	Trp	Thr	Arg	
		75					80					85				
GGC	ACC	ACT	GAA	TCC	ATC	AAC	ATG	GTG	GCA	CAA	TGC	TAT	GCG	CGT	CCG	2093
Gly	Thr	Thr	Glu	Ser	Ile	Asn	Met	Val	Ala	Gln	Cys	Tyr	Ala	Arg	Pro	
	90					95					100					
CGT	CTG	CAA	CCG	GGC	GAT	GAG	ATT	ATT	GTC	AGC	GTG	GCA	GAA	CAC	CAC	2141
Arg	Leu	Gln	Pro	Gly	Asp	Glu	Ile	Ile	Val	Ser	Val	Ala	Glu	His	His	
105				110						115					120	
GCC	AAC	CTC	GTC	CCC	TGG	CTG	ATG	GTC	GCC	CAA	CAA	ACT	GGA	GCC	AAA	2189
Ala	Asn	Leu	Val	Pro	Trp	Leu	Met	Val	Ala	Gln	Gln	Thr	Gly	Ala	Lys	
				125					130					135		
GTG	GTG	AAA	TTG	CCG	CTT	AAT	GCG	CAG	CGA	CTG	CCG	GAT	GTC	GAT	TTG	2237
Val	Val	Lys	Leu	Pro	Leu	Asn	Ala	Gln	Arg	Leu	Pro	Asp	Val	Asp	Leu	
			140					145					150			
TTG	CCA	GAA	CTG	ATT	ACT	CCC	CGT	AGT	CGG	ATT	CTG	GCG	TTG	GGT	CAG	2285
Leu	Pro	Glu	Leu	Ile	Thr	Pro	Arg	Ser	Arg	Ile	Leu	Ala	Leu	Gly	Gln	
		155					160					165				
ATG	TCG	AAC	GTT	ACT	GGC	GGT	TGC	CCG	GAT	CTG	GCG	CGA	GCG	ATT	ACC	2333
Met	Ser	Asn	Val	Thr	Gly	Gly	Cys	Pro	Asp	Leu	Ala	Arg	Ala	Ile	Thr	
	170					175					180					
TTT	GCT	CAT	TCA	GCC	GGG	ATG	GTG	GTG	ATG	GTT	GAT	GGT	GCT	CAG	GGG	2381
Phe	Ala	His	Ser	Ala	Gly	Met	Val	Val	Met	Val	Asp	Gly	Ala	Gln	Gly	
185				190						195					200	
GCA	GTG	CAT	TTC	CCC	GCG	GAT	GTT	CAG	CAA	CTG	GAT	ATT	GAT	TTC	TAT	2429
Ala	Val	His	Phe	Pro	Ala	Asp	Val	Gln	Gln	Leu	Asp	Ile	Asp	Phe	Tyr	
				205				210						215		

GCT TTT TCA GGT CAC AAA CTG TAT GGC CCG ACA GGT ATC GGC GTG CTG	2477
Ala Phe Ser Gly His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Val Leu	
220 225 230	
TAT GGT AAA TCA GAA CTG CTG GAG GCG ATG TCG CCC TGG CTG GGC GGC	2525
Tyr Gly Lys Ser Glu Leu Leu Glu Ala Met Ser Pro Trp Leu Gly Gly	
235 240 245	
GGC AAA ATG GTT CAC GAA GTG AGT TTT GAC GGC TTC ACG ACT CAA TCT	2573
Gly Lys Met Val His Glu Val Ser Phe Asp Gly Phe Thr Thr Gln Ser	
250 255 260	
GCG CCG TGG AAA CTG GAA GCT GGA ACG CCA AAT GTC GCT GGT GTC ATA	2621
Ala Pro Trp Lys Leu Glu Ala Gly Thr Pro Asn Val Ala Gly Val Ile	
265 270 275 280	
GGA TTA AGC GCG GCG CTG GAA TGG CTG GCA GAT TAC GAT ATC AAC CAG	2669
Gly Leu Ser Ala Ala Leu Glu Trp Leu Ala Asp Tyr Asp Ile Asn Gln	
285 290 295	
GCC GAA AGC TGG AGC CGT AGC TTA GCA ACG CTG GCG GAA GAT GCG CTG	2717
Ala Glu Ser Trp Ser Arg Ser Leu Ala Thr Leu Ala Glu Asp Ala Leu	
300 305 310	
GCG AAA CGT CCC GGC TTT CGT TCA TTC CGC TGC CAG GAT TCC AGC CTG	2765
Ala Lys Arg Pro Gly Phe Arg Ser Phe Arg Cys Gln Asp Ser Ser Leu	
315 320 325	
CTG GCC TTT GAT TTT GCT GGC GTT CAT CAT AGC GAT ATG GTG ACG CTG	2813
Leu Ala Phe Asp Phe Ala Gly Val His His Ser Asp Met Val Thr Leu	
330 335 340	
CTG GCG GAG TAC GGT ATT GCC CTG CGG GCC GGG CAG CAT TGC GCT CAG	2861
Leu Ala Glu Tyr Gly Ile Ala Leu Arg Ala Gly Gln His Cys Ala Gln	
345 350 355 360	
CCG CTA CTG GCA GAA TTA GGC GTA ACC GGC ACA CTG CGC GCC TCT TTT	2909
Pro Leu Leu Ala Glu Leu Gly Val Thr Gly Thr Leu Arg Ala Ser Phe	
365 370 375	
GCG CCA TAT AAT ACA AAG AGT GAT GTG GAT GCG CTG GTG AAT GCC GTT	2957
Ala Pro Tyr Asn Thr Lys Ser Asp Val Asp Ala Leu Val Asn Ala Val	
380 385 390	
GAC CGC GCG CTG GAA TTA TTG GTG GAT TAAACGCGTG CTAGAGGCAT	3004
Asp Arg Ala Leu Glu Leu Leu Val Asp	
395 400	
CAAATAAAAC GAAAGGCTCA GTCGAAAGAC TGGGCCTTTC GTTTTATCTG TTGTTTGTCTG	3064

GTGAACGCTC	TCCTGAGTAG	GACAAATCCG	CCGCCCTAGA	CCTAGGGGAT	ATATTCCGCT	3124
TCCTCGCTCA	CTGACTCGCT	ACGCTCGGTC	GTTCTGACTGC	GGCGAGCGGA	AATGGCTTAC	3184
GAACGGGGCG	GAGATTTCTC	GGAAGATGCC	AGGAAGATAC	TTAACAGGGA	AGTGAGAGGG	3244
CCGCGGCAAA	GCCGTTTTTC	CATAGGCTCC	GCCCCCTGA	CAAGCATCAC	GAAATCTGAC	3304
GCTCAAATCA	GTGGTGGCGA	AACCCGACAG	GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	3364
GCGGCTCCCT	CGTGCGCTCT	CCTGTTCTCT	CCTTTCGGTT	TACCGGTGTC	ATTCCGCTGT	3424
TATGGCCGCG	TTTGTCTCAT	TCCACGCCTG	ACACTCAGTT	CCGGGTAGGC	AGTTCGCTCC	3484
AAGCTGGACT	GTATGCACGA	ACCCCCGTT	CAGTCCGACC	GCTGCGCCTT	ATCCGGTAAC	3544
TATCGTCTTG	AGTCCAACCC	GGAAAGACAT	GCAAAGACAC	CACTGGCAGC	AGCCACTGGT	3604
AATTGATTTA	GAGGAGTTAG	TCTTGAAGTC	ATGCGCCGGT	TAAGGCTAAA	CTGAAAGGAC	3664
AAGTTTTGGT	GAATGCGCTC	CTCCAAGCCA	GTTACCTCGG	TTCAAAGAGT	TGGTAGCTCA	3724
GAGAACCTTC	GAAAAACCGC	CCTGCAAGGC	GGTTTTTTTCG	TTTTCAGAGC	AAGAGATTAC	3784
GCGCAGACCA	AAACGATCTC	AAGAAGATCA	TCTTATTAAT	CAGATAAAAT	ATTTCTAGAT	3844
TTCAGTGCAA	TTTATCTCTT	CAAATGTAGC	ACCTGAAGTC	AGCCCCATAC	GATATAAGTT	3904
GTTACTAGTG	CTTGGATTCT	CACCAATAAA	AAACGCCCCG	CGGCAACCGA	GCGTTCTGAA	3964
CAAATCCAGA	TGGAGTTCTG	AGGTCATTAC	TGGATCTATC	AACAGGAGTC	CAAGCGAGCT	4024
CTCGAACCCC	AGAGTCCCGC	TCAGAAGAAC	TCGTCAAGAA	GGCGATAGAA	GGCGATGCGC	4084
TGCGAATCGG	GAGCGGCGAT	ACCGTAAAGC	ACGAGGAAGC	GGTCAGCCCA	TTGCGCGCCA	4144
AGCTCTTCAG	CAATATCACG	GGTAGCCAAC	GCTATGTCCT	GATAGCGGTC	CGCCACACCC	4204
AGCCGGCCAC	AGTCGATGAA	TCCAGAAAAG	CGGCCATTTT	CCACCATGAT	ATTCGGCAAG	4264
CAGGCATCGC	CATGGGTCAC	GACGAGATCC	TCGCCGTCGG	GCATGCGCGC	CTTGAGCCTG	4324
GCGAACAGTT	CGGCTGGCGC	GAGCCCCTGA	TGCTCTTCGT	CCAGATCATC	CTGATCGACA	4384
AGACCGGCTT	CCATCCGAGT	ACGTGCTCGC	TCGATGCGAT	GTTTCGCTTG	GTGGTCGAAT	4444
GGGCAGGTAG	CCGGATCAAG	CGTATGCAGC	CGCCGCATTG	CATCAGCCAT	GATGGATACT	4504
TTCTCGGCAG	GAGCAAGGTG	AGATGACAGG	AGATCCTGCC	CCGGCACTTC	GCCCAATAGC	4564

```

AGCCAGTCCC TTCCCGCTTC AGTGACAACG TCGAGCACAG CTGCGCAAGG AACGCCCCGTC      4624
GTGGCCAGCC ACGATAGCCG CGCTGCCTCG TCCTGCAGTT CATTCAGGGC ACCGGACAGG      4684
TCGGTCTTGA CAAAAAGAAC CGGGCGCCCC TCGCTGACA GCCGGAACAC GGCGGCATCA      4744
GAGCAGCCGA TTGTCTGTTG TGCCCAGTCA TAGCCGAATA GCCTCTCCAC CCAAGCGGCC      4804
GGAGAACCTG CGTGCAATCC ATCTTGTTCA ATCATGCGAA ACGATCCTCA TCCTGTCTCT      4864
TGATCAGATC TTGATCCCCT GCGCCATCAG ATCCTTGGCG GCAAGAAAGC CATCCAGTTT      4924
ACTTTGCAGG GCTTCCCAAC CTTACCAGAG GGCGCCCCAG CTGGCAATTC C      4975

```

## (2) INFORMATION FOR SEQ ID No: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 384 Amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID No: 14:

```

Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
 1              5              10              15
Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
      20              25              30
Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
      35              40              45
Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
      50              55              60
Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
      65              70              75              80
Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
      85              90              95
Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
      100              105              110
Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
      115              120              125

```

```

Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
 130                               135                               140

Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
 145                               150                               155                               160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
                               165                               170                               175

Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
                               180                               185                               190

Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
 195                               200                               205

Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
 210                               215                               220

Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
 225                               230                               235                               240

Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
                               245                               250                               255

Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
 260                               265                               270

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
 275                               280                               285

Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
 290                               295                               300

Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
 305                               310                               315                               320

Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
                               325                               330                               335

Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
 340                               345                               350

Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
 355                               360                               365

Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
 370                               375                               380

```

(2) INFORMATION FOR SEQ ID No: 15:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 401 Amino acids

(B) TYPE: Amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID No: 15:

Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu  
 1 5 10 15

Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro  
 20 25 30

Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly  
 35 40 45

Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg  
 50 55 60

Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp  
 65 70 75 80

Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met  
 85 90 95

Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile  
 100 105 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met  
 115 120 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala  
 130 135 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg  
 145 150 155 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys  
 165 170 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val  
 180 185 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val  
 195 200 205

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr  
 210 215 220

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu  
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser  
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly  
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp  
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu  
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser  
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val  
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu  
340 345 350

Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val  
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp  
370 375 380

Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val  
385 390 395 400

Asp